Clustering Of Ecg Using D-Stream Algorithm

Vaishali Yeole
Department of computer Engg.
K.C college of Engg,
Thane (E).

Jyoti Kadam
Department of computer Engg.
K.C college of Engg
Thane (E).

Abstract──**The overall objective of this paper is to design and implement Clustering of ECG Using D-Stream algorithm which uses the Dstream clustering algorithm to find out the heart disease with respect to ECG. We implemented a system which converts the ECG of patient into the cluster and later on map the cluster with the existing cluster in the database. Successful implementation of the final system would be of benefit to all involved in the use of electrocardiography as access to, and movement of, the patient would not be impeded by the physical constraints imposed by the cables. Most aspects of the design would also be portable to applications, making the work relevant to a vast range of systems where movement of sensors is desirable and constrained by hard-wired links.**

The design and implementation of the application is based on data stream algorithm where it read the "n" no. of inputs once and maps with the existing cluster into the database, if the match is found then the respective heart disease will detect, if no then it will add into the database.

Keywords── Density grid, clustering, data streams, ECG.

1. INTRODUCTION

In the data stream scenario, input arrives very rapidly and there is limited memory to store the input. Algorithms have to work with one or few passes over the data, space less than linear in the input size or time significantly less than the input size. In the past few years, a new theory has emerged for reasoning about algorithms that work within these constraints on space, time, and number of passes. Some of the methods rely on metric embeddings, pseudo-random computations, and sparse approximation theory and communication complexity.

Density-based clustering has been long proposed as another major [5] clustering algorithm. We find the density based method a natural and attractive basic clustering algorithm for data streams, because it can find arbitrarily shaped clusters, it can handle noises and is a one-scan algorithm that needs to examine the raw data only once. Further, it does not demand a prior knowledge of the number of clusters k as the k-means algorithm does.[2]
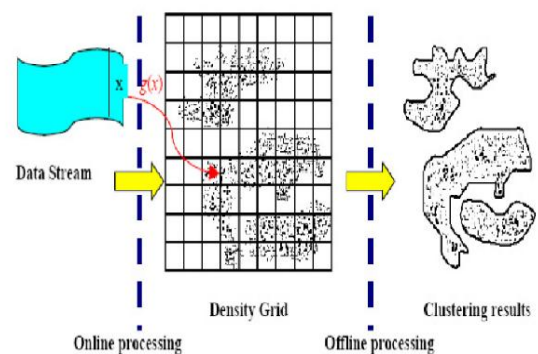

.

Figure 1.Illustration of the use of density grid.

In this, we propose D-Stream, a density-based clustering framework for data streams. It is not a simple switch-over to use density-based instead of k-means algorithms for data streams. There are two main technical challenges. First, it is not desirable to

treat the data stream as a long sequence of static data since we are interested in the evolving temporal feature of the data stream.

To capture the dynamic changing of clusters, we propose an innovative scheme that associates a decay factor to the density of each data point. Unlike the CluStream architecture which asks the users to input the target time duration for clustering, the decay factor provides a novel mechanism for the system to dynamically and automatically form the clusters by placing more weights on the most recent data without totally discarding the historical information.

In addition, D-Stream does not require the user to specify the number of clusters k. Thus, D-Stream is particularly suitable for users with little domain knowledge on the application data. Second, due to the large volume of stream data, it is impossible to retain the density information for every data record. Therefore, we propose to partition the data space into discretized fine grids and map new data records into the corresponding grid [1] .

Thus, we do not need to retain the raw data and only need to operate on the grids. However, for high-dimensional data, the number of grids can be large. Therefore, how to handle with high dimensionality and improve scalability is a critical issue. Fortunately, in practice, most grids are empty or only contain few records and a memory-efficient technique for managing such a sparse grid space is developed in D-Stream.

2. FEATURE EXTRACTION OF ECG

The electrical signals described are measured by the ECG where each heart beat is displayed as a series of electrical waves characterized by peaks and valleys. An ECG gives two major kinds of information. First, by measuring time intervals on the ECG, the duration of the electrical wave crossing the heart can be determined and consequently we can determine whether the electrical activity is normal or slow, fast or irregular. [4]Second, by measuring the amount of electrical activity passing through the heart muscle, a pediatric cardiologist may be able to find out if parts of the heart are too large or are overworked. The frequency range of an ECG signal is [0.05-100] Hz and its dynamic range is [1-10] mV. The ECG signal is characterized by five peaks and valleys labeled by successive letters of the alphabet P, Q, R, S and T. A good performance of an ECG analyzing system depends heavily upon the accurate and reliable detection of the QRS complex, as well as the T and P waves. The P wave represents the activation of the upper chambers of the heart, the atria while the QRS wave (or complex) and T wave represent the excitation of the ventricles or the lower chambers of the heart. The detection of the QRS complex is the most important task in automatic ECG signal analysis.

3. D-STREAM ALGORITHM

The D-stream algorithm is explained as follows [4]

1. procedure D-Stream

2. Tc = 0;

3. Initialize an empty hash table grid list;

4. while data stream is active do

5. read record x = (x1, x2, · · · , xd);

6. determine the density grid g that contains x;

7. if(g not in grid list) insert g to grid list;

8. update the characteristic vector of g;

9. if tc == gap then

10. call initial clustering(grid list);

11. end if

12. if tc mod gap == 0 then

13. detect and remove sporadic grids from grid list;

14. call adjust clustering(grid list);

15. end if

16. tc = tc + 1;

17. end while

18. end procedure
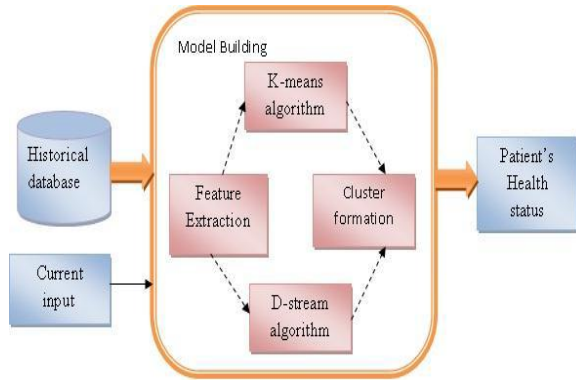
4. OVERVIEW OF DSTREAM

The procedure adjust clustering () is executed every gap time steps. Therefore, adjust clustering () will not affect the overall efficiency of the algorithm[4].

Running time is less than the time needed for the online component to process gap data records.

We only consider those grids that are maintained in grid list. Therefore, although the number of possible grids is huge for high-dimensional data, most empty or infrequent grids are discarded, which saves computing time and makes our algorithm very fast without deteriorating clustering quality.

We introduce a new concept on the attraction between grids in order to improve the quality of density-based clustering. Further, a sophisticated and theoretically sound technique is developed to detect and remove the sporadic grids in order to dramatically improve the space and time efficiency without affecting the clustering results.

The technique makes high-speed data stream clustering feasible without degrading the clustering quality.

5. RESULT ANALYSIS
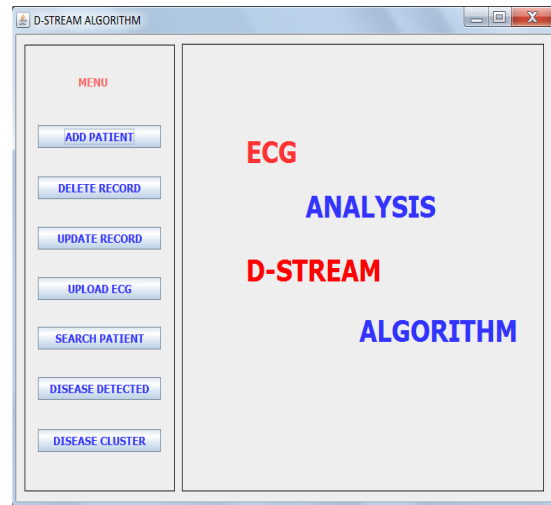
Figure 5.1 shows main screen of result

Figure 5.1 Mainframe Screen

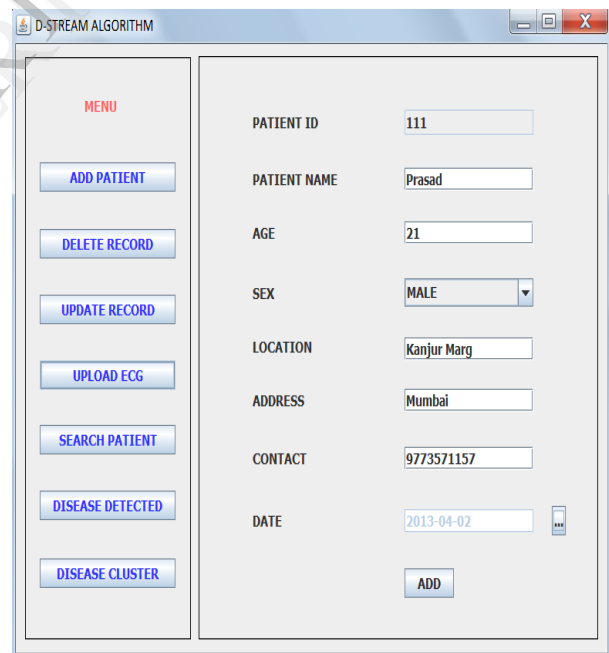Figure 5.2 takes complete information about new patient and record stored into a system's database.

Figure 5.2 To add a new record
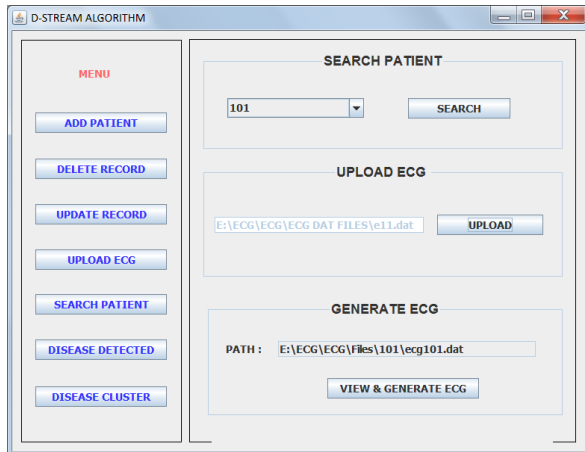
Figure 5.3 use to upload ECG of patient .
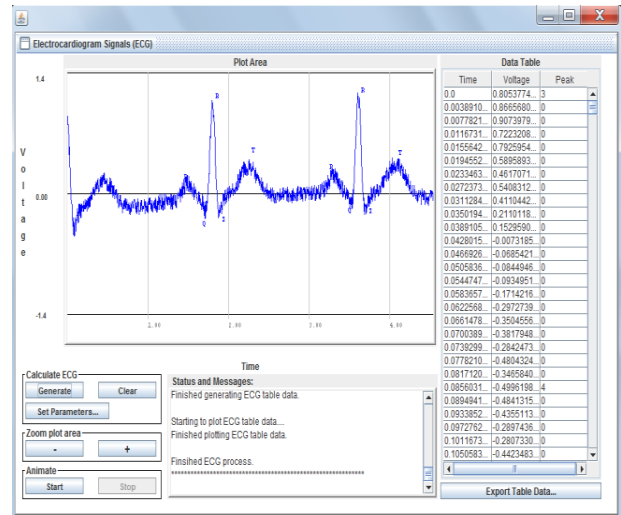
Figure 5.3 Upload ECG of a patient

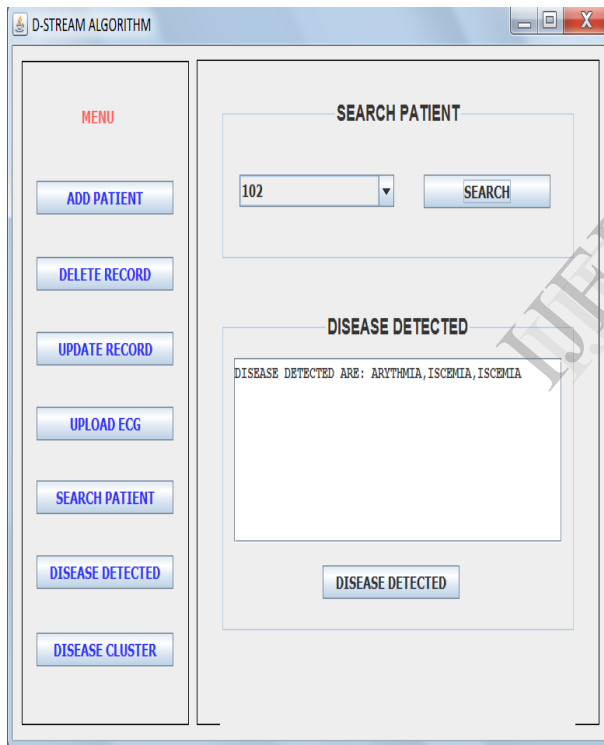figure 5.4 search the patient's record and detect number of disease by which the patient is suffering.

Figure 5.4 Disease Detected

Figure 5.5 shows the result in form of graph of uploaded ECG

Figure 5.5.: Graph of uploaded ecg

Table1: Comparison of Kmeans and

Dstream algorithms

PARAMETER	DSTREAM	K-MEANS
Efficiency	More	Less
Memory Requirement	Less	More
Complexity	Difficult To Implement	Easy To Implement
Cluster	It Can Identify Any Type Of Arbitrary Shape Cluster	It Identifies Only Spherical Shape Cluster
Grid	Maintains Grid For A Group Of Cells	No Concept Of Grid Is Present
Data Knowledge	Does Not Need Prior Knowledge Of Cluster	Needs prior Knowledge Of The Cluster
Noise	Can Handle Noise And Outliers	Cannot Handle Noise And Outliers

6. CONCLUSION

In this paper, we propose D-Stream, a new framework for clustering stream data. The algorithm maps each input data into a grid, computes the density of each grid, and clusters the grids using a density-based algorithm. In contrast to previous algorithms based on k-means, the proposed algorithm can find clusters of arbitrary shapes. The algorithm also proposes a density decaying scheme that can effectively adjust the clusters in real time and capture the evolving behaviors of the data stream [6]. We introduce a new concept on the attraction between grids in order to improve the quality of density-based clustering. Further, a sophisticated and theoretically sound technique is developed to detect and remove the sporadic grids in order to dramatically improve the space and time efficiency without affecting the clustering results. The technique makes high-speed data stream clustering feasible without degrading the clustering quality.

7. REFERENCES

[1] J. Sander, M. Ester, H. Kriegel, and X. Xu. Density-based clustering in spatial databases: The algorithm gdbscan and its applications. Data Min. Knowl. Discov., 2(2):169–194, 1998.

[**2**] We have also referred ECG site for database "The MIT-BIH Arrhythmia Database," http://physionet.ph.biu.ac.il/physiobank/database /mitdb/

[3] V.X. Afonso, W.J. Tompkins, T.Q. Nguyen, and S. Luo, "ECG beat detection using filter banks" , IEEE Trans. Biomed. Eng., vols. 46, pp.192–202, 1999.

[4] J. Beringer and E. H¨ullermeier. Online-clustering of parallel data streams. *Data and Knowledge Engineering*, 58(2):180–204, 2006.

[5] C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu. A framework for clustering evolving data streams. In *Proc.VLDB*, pages 81–92, 2003.

[6] http:// www.cse/msu.edu/~ptan/papers.html